YEAST VOL. **13:** 639–645 (1997)

Sequence Comparison of the Ty1 and Ty2 Elements of the Yeast Genome Supports the Structural Model of the $tRNA_i^{Met}$–Ty1 RNA Reverse Transcription Initiation Complex

S. FRIANT[1], T. HEYMAN[2], O. POCH[3], M. WILHELM[1] AND F. X. WILHELM[1]*

[1]*Unité Propre de Recherche 9002 du Centre National de la Recherche Scientifique, Institut de Biologie Moléculaire et Cellulaire, 15 rue René Descartes, 67084 Strasbourg Cedex, France*
[2]*Unité Mixte de Recherche 216 du Centre National de la Recherche Scientifique, Institut Curie Biologie, Centre Universitaire, 91405 Orsay Cedex, France*
[3]*Unité Propre de Recherche 9005 du Centre National de la Recherche Scientifique, Institut de Biologie Moléculaire et Cellulaire, 15 rue René Descartes, 67084 Strasbourg Cedex, France*

Received 1 November 1996; accepted 15 January 1997

In the reverse transcription initiation complex of the yeast Ty1 retrotransposon, interaction between the template RNA and primer $tRNA_i^{Met}$ is not limited to base pairing of the primer binding site (PBS) with ten nucleotides at the 3′ end of $tRNA_i^{Met}$, but three regions named boxes 0, 1 and 2.1 interact with the T and D stems and loops of $tRNA_i^{Met}$. Sequence comparison of 33 Tyl elements and 13 closely related Ty2 elements found in the yeast genome shows that the nucleotide sequence of all elements is highly conserved in the region spanning the PBS and the three boxes. Since this domain of the template RNA encodes a portion of protein TyA, we have calculated its amino acid profile and its nucleotide profile to evaluate the role played by nucleotide sequence conservation in the selection for TyA function and in the maintenance of base pairing interactions for the priming function of Tyl RNA. Our results show that the nucleotide sequence conservation of Ty1 RNA is constrained not only by selection for Ty1 function but also by maintenance of a given nucleotide sequence able to base pair with the $tRNA_i^{Met}$ in the primer–template initiation complex. © 1997 by John Wiley & Sons, Ltd.

Yeast 13: 639–645, 1997.
No. of Figures: 3. No. of Tables: 0. No. of References: 10.

KEY WORDS — Ty1 retrotransposon; Ty2; primer binding site

Reverse transcription of the yeast Ty1 retrotransposon is primed by the initiator methionine tRNA ($tRNA_i^{Met}$) base paired to the minus-strand primer binding site (PBS) near the 5′ end of the Ty1 genomic RNA. To understand the molecular basis of the primer–template interaction we have recently made a structural probing study of the $tRNA_i^{Met}$–Ty1 RNA binary complex (Friant *et al.*, 1996). We have found that interaction between the primer $tRNA_i^{Met}$ and Ty1 RNA is not limited to base pairing of ten nucleotides at the 3′ end of $tRNA_i^{Met}$ with the PBS of Ty1 RNA but that three regions named boxes 0, 1 and 2.1 interact with the T and D stems and loops of $tRNA_i^{Met}$ (Figure 1). We also showed that another region of Ty1 RNA named box 2.2, which has the same sequence as box 2.1 and could potentially interact with the $tRNA_i^{Met}$ does not interact with the $tRNA_i^{Met}$ in the $tRNA_i^{Met}$–Ty1 RNA complex.

Formation of the complex induces important structural rearrangements in the Ty1 RNA domain spanning the PBS, box 0, box 1 and box 2.1 with the exception of a stem-loop immediately 3′ of box 0 (stem-loop A in Figure 1). The structure of the region 5′ of the PBS (nucleotides 70–94) and 3′ of

*Correspondence to: F. X. Wilhelm.
Contract grant sponsor: Association pour la Recherche contre le Cancer.
Contract grant sponsor: Ligue Nationale contre le Cancer.

CCC 0749–503X/97/070639–07 $17.50

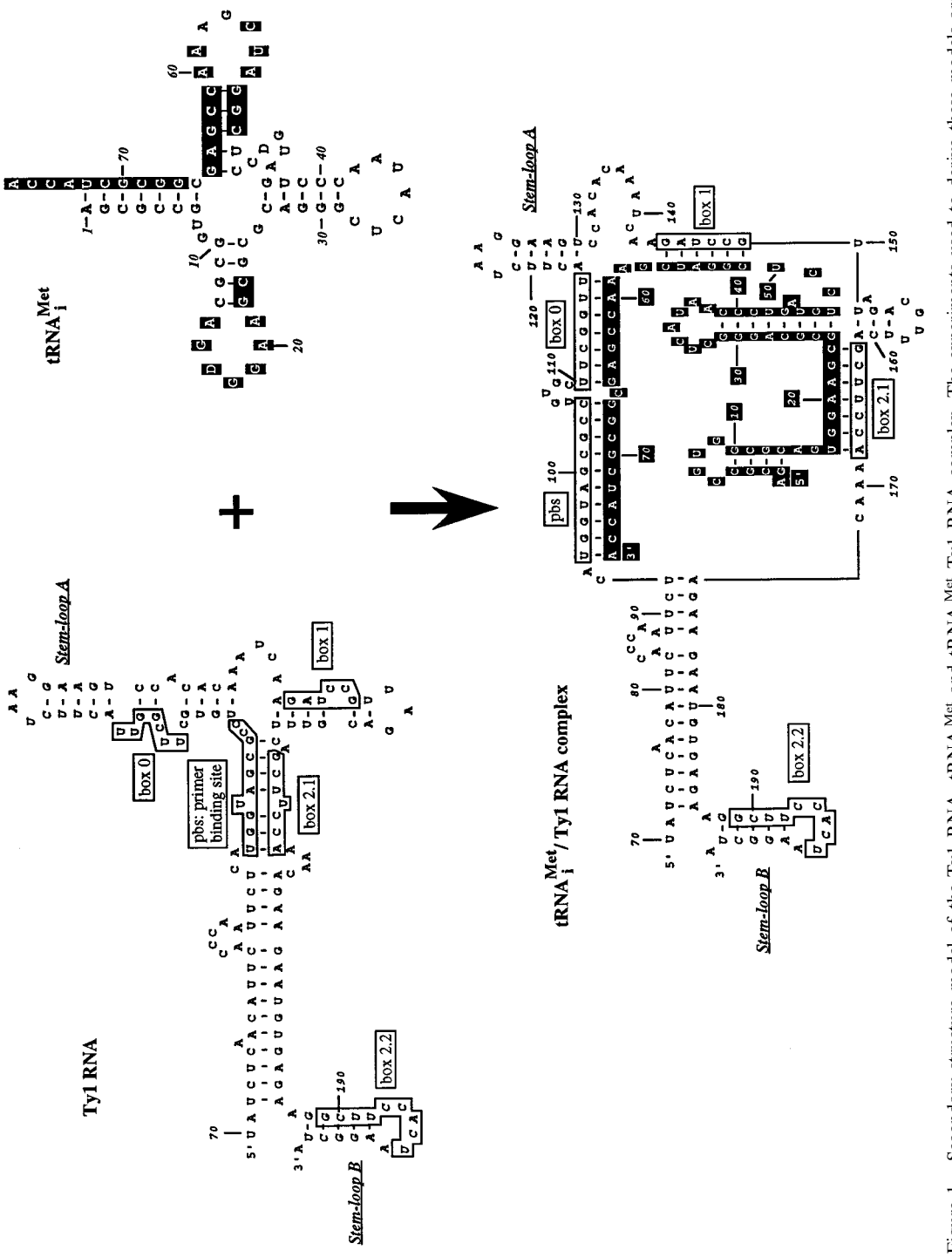


Figure 1. Secondary structure model of the Ty1 RNA, tRNA$_i^{Met}$ and tRNA$_i^{Met}$–Ty1 RNA complex. The experiments used to derive these models are described in Friant *et al.* (1996). The sequence of the Ty1 RNA is that of the Ty1-H3 element cloned by Boeke *et al.* (1988).

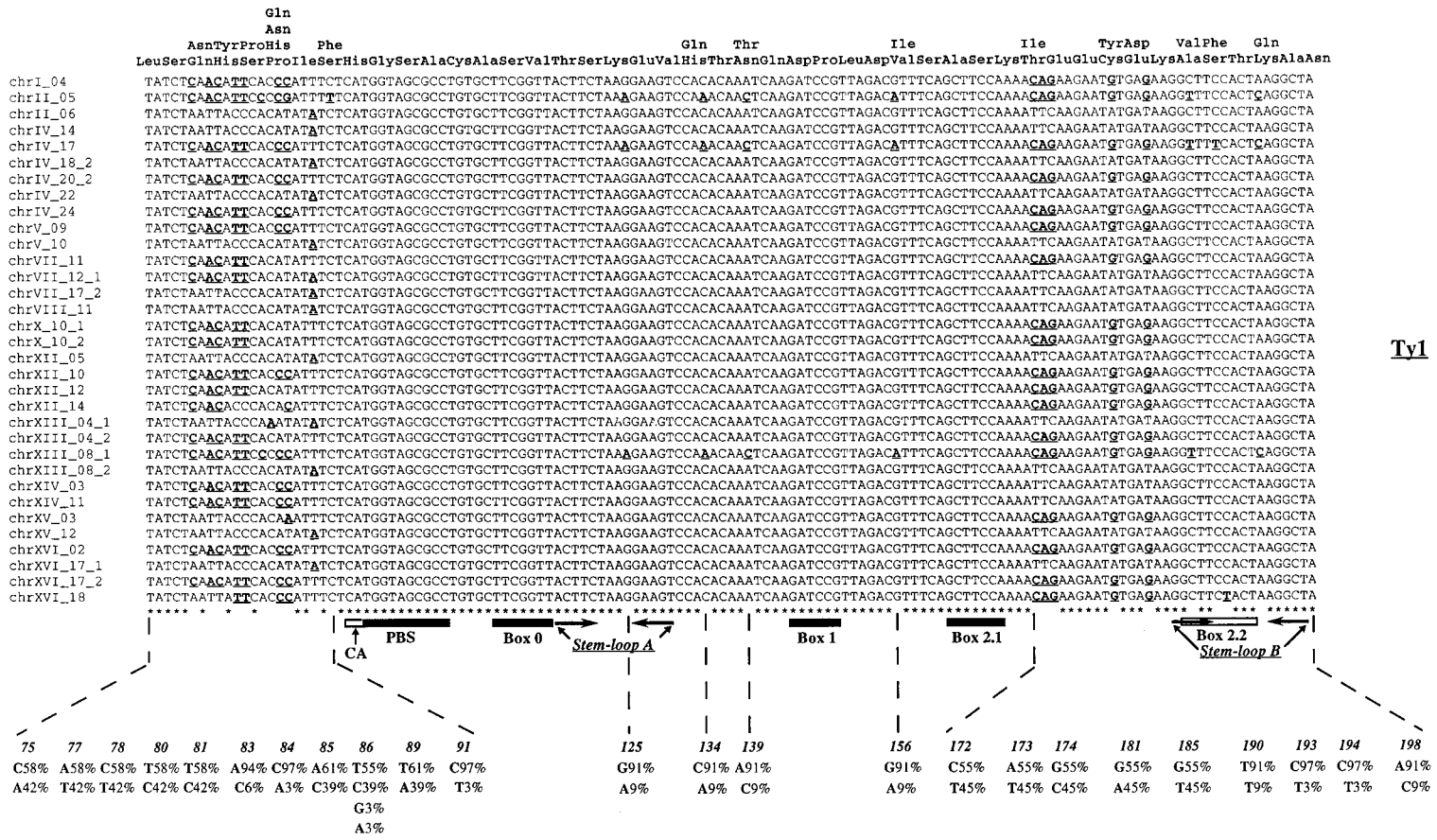
box 2.1 (nucleotides 169–204) was not rearranged upon formation of the complex. The importance of the boxes for the function of the retrotransposon was also inferred from *in vivo* experiments showing that mutations disrupting some complementary base pairs between the boxes and the primer tRNA impair transposition of the Ty1 element (Keeney *et al.*, 1995; Wilhelm *et al.*, 1994). Sequence and structural studies of the 5′ region of retroviral RNAs indicate that extended contacts of the primer tRNA with sequences outside the PBS exist in several retroviruses (Isel *et al.*, 1995; Leis *et al.*, 1993) and have a functional role (Cordell *et al.*, 1979; Haseltine *et al.*, 1976). It has been proposed that interactions between seven bases of the T loop of the $tRNA^{Trp}$ primer and sequences in the U5 region of the 5′ LTR may be required in an early step of reverse transcription by Rous sarcoma virus reverse transcriptase and may be necessary for annealing of the primer tRNA to the viral RNA.

In our previous study (Friant *et al.*, 1996) the importance of the boxes and of extended interactions between the primer $tRNA_i^{Met}$ and Ty1 RNA was supported by a sequence comparison of the Ty1 and closely related Ty2 elements. We showed that the primary sequence of the boxes was strictly conserved in 14 elements found in the EMBL GenBank. We have now made a computer search of Ty1 and Ty2 elements in the recently published sequence of the yeast genome. The PBS sequence, TGGTAGCGCC, was used to screen the yeast genome for Ty1 and Ty2 elements. Forty-six elements (33 Ty1 elements and 13 Ty2 elements) with an exact copy of the PBS sequence were found. This number is in good agreement with the estimated number of Ty1 and Ty2 elements in the haploid genome of *Saccharomyces cerevisiae*, i.e. 25–35 copies of Ty1 and 10–15 copies of Ty2 (Boeke and Sandmeyer, 1991). Another hallmark of the Ty1 and Ty2 elements is the PPT (polypurine tract) sequence, TGGGT GGTA, which serves as a primer to initiate synthesis of plus-strand DNA (Heyman *et al.*, 1995). We have used this sequence to make a second screen of the yeast genome for Ty1 and Ty2 elements. Forty-five of the elements found in the first screen were found in the second screen. The 46th element, Ty1 chromosome IV_17 is truncated since no PPT sequence could be found in the 68916 nucleotides of chromosome IV between the PBS of this element and the PBS of the next element on chromosome IV (Ty1 chromosome IV_18_1). In most Ty1 elements (see Figure 2 for exceptions) two exact copies of the PPT sequence located 3447–3450 and 5241–5244 nucleotides downstream of PBS were found. These two PPTs probably correspond to the two functional PPTs recently identified in the genome of the Ty1-H3 element (Heyman *et al.*, 1995). In ten Ty2 elements, four exact copies of the PPT sequence located 2970, 3126, 3492 and 5286 nucleotides downstream of the PBS were found. It is not known whether all PPT sequences are functional in Ty2.

The sequences of 33 Ty1 and 13 Ty2 elements were aligned using the Clustal multiple alignment program. Nucleotide sequence comparison of the RNA domain (nucleotides 70–204) used in our previous structural study (Friant *et al.*, 1996) is shown in Figure 2A (Ty1) and 2B (Ty2). It is striking that the domain including the PBS and the three boxes 0, 1 and 2.1 (nucleotides 94–169), which has been shown to interact with the primer $tRNA_i^{Met}$ in the model of $tRNA_i^{Met}$–Ty1 RNA reverse transcription initiation complex, is highly conserved in all Ty1 and Ty2 elements: in this domain only three Ty1 elements have variable nucleotides at positions 125, 134, 139 and 156 and only four mutations are observed in Ty2 elements.

The nucleotide sequence of the Ty2 elements is more variable in the regions 5′ of the PBS (nucleotides 70–94). For Ty1 elements the nucleotide sequence is more variable in both the regions 5′ of the PBS (nucleotides 70–94) and 3′ of box 2.1 (nucleotides 169–204). The extent of heterogeneity, given by the number value at each variable position (Figure 2), is also higher in these regions. Interestingly, the sequence of box 2.2 (nucleotides 189–197), which does not interact with $tRNA_i^{Met}$ in the structural model presented in Figure 1, is not strictly conserved in the Ty1 elements; a strict conservation of its sequence is probably not necessary for the formation of a stable $tRNA_i^{Met}$–Ty1 RNA complex.

The RNA region spanning nucleotides 70 to 204 encodes a portion of protein TyA (TyA is equivalent to the *gag* protein of retroviruses) whose amino acid sequence is indicated above the nucleotide sequence in Figure 2. It is obvious that constraints to maintain the reading frame of the protein will also constrain the nucleotide sequence from mutating. To evaluate the role played by nucleotide sequence conservation in the selection for TyA function and in the maintenance of base pairing interactions for the priming function of



Ty1

Figure 2. (A) Sequence alignment of nucleotides 70 to 204 of Ty1 elements. A Ty1 RNA fragment spanning these nucleotides was used to derive the structural model shown in Figure 1. The PBS and the boxes are underlined. Stem-loops A and B are indicated. The dinucleotide CA important for integration of the element into the host genome is indicated. The variable nucleotides are indicated in bold. The extent of variation is given at each variable position. The amino acid sequence is indicated above the nucleotide sequence. The PBS sequence, TGGTAGCGCC, was used to screen the yeast genome for Ty1 elements. Thirty-three Ty1 elements with an exact copy of the PBS sequence and two PPT sequences located 3447–3450 and 5241–5244 nucleotides downstream of PBS were found with the following exceptions: the Ty1 chromosome IV_17 element is truncated. In Ty1 chromosome XIV_11, the PPT sequences are located 3429 and 5223 nucleotides downstream of the PBS. In Ty1 chromosome VIII_11, the PPT sequences are located 3560 and 5357 nucleotides downstream of the PBS. Ty1 chromosome IV_20_2 has only one PPT sequence located 4815 nucleotides downstream of the PBS.

Figure 2. (B) Sequence alignment of nucleotides 70 to 204 of Ty2 elements. There is a gap in the alignment of Ty2 chromosome XIV_12 has one extra T residue which might have arisen from a sequencing error. The coding sequence of Ty2 chromosome VII_12_2 is interrupted by a TAA ochre stop codon. In ten of the 13 Ty2 elements, four exact copies of the PPT sequence located 2970, 3126, 3492 and 5286 nucleotides downstream of the PBS were found. In Ty2 chromosome XIV_12 the PPT sequence 2970 nucleotides downstream of the PBS is missing. Ty2 chromosome XII_20 has only one PPT sequence located 4769 nucleotides downstream of the PBS. Ty2 chromosome III_02 has only two PPT sequences located 3495 and 5289 nucleotides downstream of the PBS.

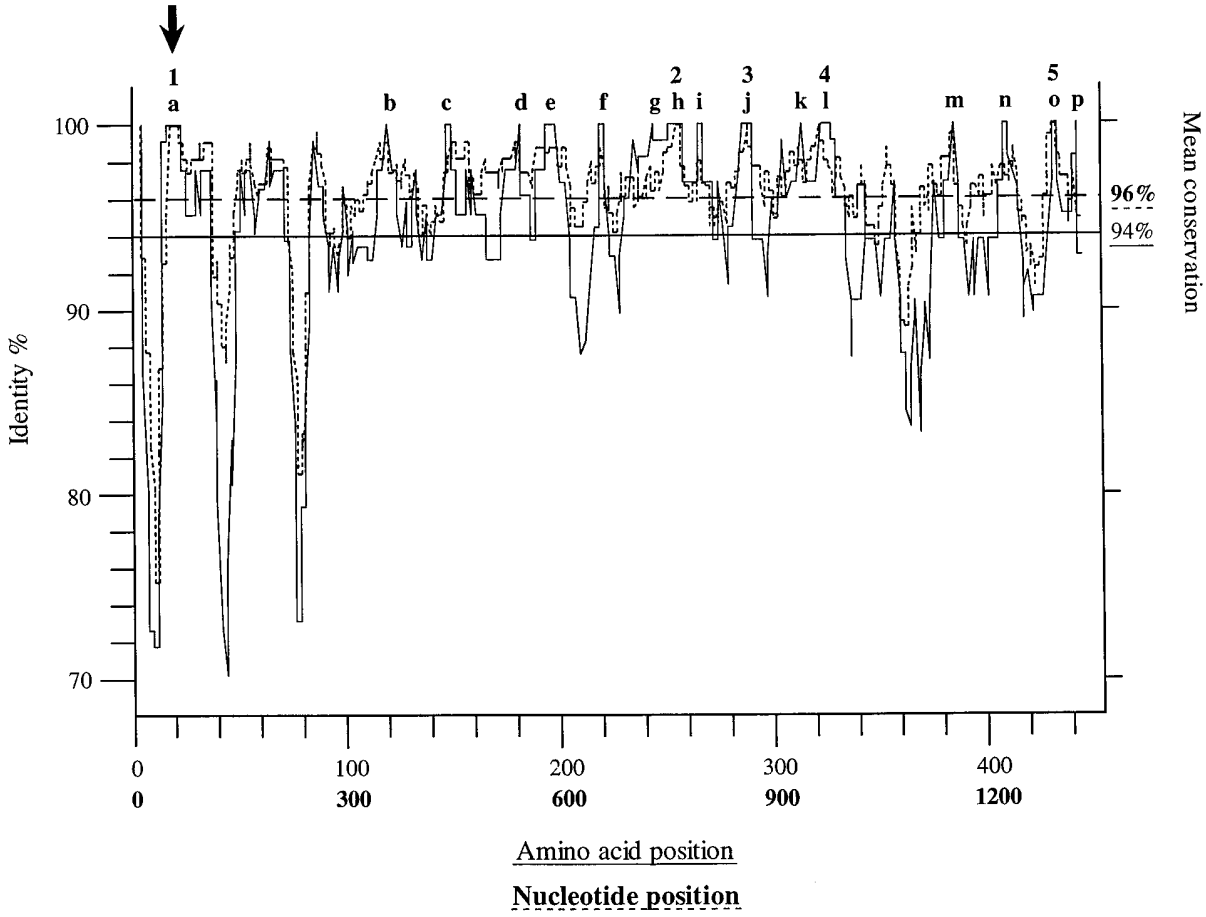


Figure 3. Identity profiles obtained after nucleotide (– – –) and amino acid (——) sequence alignment of the TyA open reading frame of all Ty1 elements. Position 1 of the nucleotide sequence corresponds to the A residue of the initiator ATG codon of TyA. The percentage of identity was calculated within a window of 21 nucleotides or seven amino acids running with a step of 1. Identity was scored as 100% when the nucleotide or amino acid sequence was strictly conserved in all elements. Sixteen domains (a to p) of the protein sequence are 100% identical and five domains (1 to 5) of the nucleotide sequence are 100% identical. The region of highest nucleotide conservation (domain 1) indicated by an arrow corresponds to nucleotides 92 to 124 (see Figure 2A) spanning the PBS, box 0 and the 5′ part of stem-loop A. The mean nucleotide or amino acid conservations over the entire aligned sequences are indicated.

Ty1 RNA, the amino acid and nucleotide identity profiles of the TyA open reading frame were calculated. These profiles were obtained after alignment of the sequences of the entire TyA open reading frame of all Ty1 elements.

In the amino acid plot (Figure 3, solid line), 16 domains (a to p) of the protein are 100% identical whereas in the nucleotide plot (Figure 3, dashed line) only five domains (1 to 5) are 100% identical. The broader domain of strict identity in the amino acid plot (domain h) corresponds to a region encompassing amino acids 244 to 256 encoded by nucleotides 755 to 794. In this region, a domain of strict nucleotide identity includes nucleotides 765 to 790 (domain 2). It is striking that the broader domain of strict nucleotide identity (domain 1 indicated by an arrow in Figure 3) includes the PBS and the boxes (nucleotides 65–97) involved in base pairing interactions with the $tRNA_i^{Met}$ in the reverse transcription initiation complex.

These results clearly suggest that the nucleotide sequence conservation observed in the domain spanning the PBS, box 0, box 1 and box 2.1 is constrained not only by selection for TyA function but also by maintenance of a given nucleotide sequence able to base pair with the primer

$tRNA_i^{Met}$ used to initiate reverse transcription of the Ty1 element.

The present study provides some support for extended interactions between the primer $tRNA_i^{Met}$ and template RNA of Ty1 elements. Preliminary evidence suggested that specific base pairing interactions between box 0, box 1 and box 2.1 and the T and D loops and arms of the primer $tRNA_i^{Met}$ may be required for priming of reverse transcription of the Ty1 element *in vivo* (Wilhelm *et al*., 1994). Experiments are now in progress to test the importance of base pairing interactions between each box of Ty1 RNA and the primer $tRNA_i^{Met}$ in the priming function of the Ty1 element.

ACKNOWLEDGEMENTS

We thank F. Nachon and P. Auffinger for their help with the sequence alignment program. This work was supported in part by grants from the Association pour la Recherche contre le Cancer (ARC) and from the Ligue Nationale contre le Cancer, Comité Départemental du Haut-Rhin.

REFERENCES

Boeke, J. D., Eichinger, D., Castrillon, D. and Fink, G. R. (1988). The *Saccharomyces cerevisiae* genome contains functional and nonfunctional copies of transposon Ty1. *Mol. Cell. Biol.* **8,** 1432–1442.

Boeke, J. D. and Sandmeyer, S. (1991). *Yeast Transposable Elements.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Cordell, B., Swanstrom, R., Goodman, H. M. and Bishop, J. M. (1979). $tRNA^{Trp}$ as primer for RNA-directed DNA polymerase: structural determinants and function. *J. Biol. Chem.* **254,** 1866–1874.

Friant, S., Heyman, T., Wilhelm, M. L. and Wilhelm, F. X. (1996). Extended interactions between the primer $tRNA_i^{Met}$ and genomic RNA of the yeast Ty1 retrotransposon. *Nucleic Acids Res.* **24,** 441–449.

Haseltine, W. A., Kleid, D. G., Panet, A., Rothenberg, E. and Baltimore, D. (1976). Ordered transcription of RNA tumor virus genomes. *J. Mol. Biol.* **106,** 109–131.

Heyman, T., Agoutin, B., Friant, S., Wilhelm, F. X. and Wilhelm, M. L. (1995). Plus-strand DNA synthesis of the yeast retrotransposon Ty1 is initiated at two sites, PPT1 next to the 3′ LTR and PPT2 within the pol gene. PPT1 is sufficient for Ty1 transposition. *J. Mol. Biol.* **253,** 291–303.

Isel, C., Ehresmann, C., Keith, G., Ehresmann, B. and Marquet, R. (1995). Initiation of reverse transcription of HIV-1: secondary structure of the HIV-1 RNA/tRNA Lys3 (template/primer) complex. *J. Mol. Biol.* **247,** 236–250.

Keeney, J. B., Chapman, K. B., Lauermann, V., *et al*. (1995). Multiple molecular determinants for retrotransposition in a primer tRNA. *Mol. Cell. Biol.* **15,** 217–226.

Leis, J., Aiyar, A. and Cobrinik, D. (1993). *Regulation of Initiation of Reverse Transcription of Retroviruses.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Wilhelm, M., Wilhelm, F. X., Keith, G., Agoutin, B. and Heyman, T. (1994). Yeast Ty1 retrotransposon: the minus-strand primer binding site and a cis-acting domain of the Ty1 RNA are both important for packaging of primer tRNA inside virus-like particles. *Nucleic Acids Res.* **22,** 4560–4565.

